NSB 2014 Lund 15-19 June
Proceedings of 10th Nordic Symposium on Building Physics

Reliability of meta-modelling in robust low-energy dwelling design

Liesje Van Gelder, M.Sc.
Hans Janssen, Professor
Staf Roels, Professor

Building Physics Section, KU Leuven, Belgium

KEYWORDS: *meta-modelling, probabilistic design, reliability*

SUMMARY:
In building design, energy demand and life-cycle costs are commonly calculated based on deterministic stationary or dynamic simulations. However, many contributing parameters are inherently uncertain, resulting in potentially unreliable performance predictions. To overcome this, a probabilistic design method is recommended to take uncertainties into account. Such an uncertainty-based optimisation often requires many simulations, making it extremely time-consuming. Here, meta-modelling can be of high interest. A meta-model aims to mimic the original numerical model with a simplified fast model. However, the simplicity of the meta-model - some aspects of the original model are inherently neglected - might affect the reliability of the results. This topic is investigated in this paper by means of a case study of robust cost optimisation of a low-energy dwelling. To maximise calculation efficiency, the meta-model has to be trained on as few samples as possible, taking into account meta-model reliability. Hence, a meta-modelling procedure is proposed and result reliability is investigated for meta-models based on different sample sizes. It is concluded that meta-models can be reliably used in probabilistic design and built with a reasonable sample size.

1. Introduction

Building performance optimisation typically uses deterministic simulations to select the best performing design according to one or more performance indicators, such as energy demand, life-cycle cost, … As many contributing parameters are inherently uncertain, such deterministic optimisation not necessarily leads to the best performing design. Therefore, probabilistic performance optimisation is recommended to take these uncertainties into account. The global framework of probabilistic design is described in section 1.1. Unfortunately, such a probabilistic design often requires significant simulation effort. To reduce calculation time, meta-modelling is of high interest as the original model is replaced by a fast simplified equation-based model, as described in section 1.2.

Because meta-model reliability is essential in probabilistic design, this paper investigates this aspect with a simplified robust cost optimisation of a low-energy dwelling. For that purpose, optimisation is performed on both the original model and several meta-models, differing in sample size.

The case study is described in section 2, the results are shown in section 3. The main observations and recommendations with respect to meta-model reliability are summarised in section 4.

1.1 Probabilistic design

In design problems, contributing input parameters can be divided into three categories. Design parameters, such as the thermal resistance of a wall, the type of ventilation system , … are fully controllable and their values are to be selected. Inherently uncertain parameters, such as the impact of workmanship, the actual ventilation rate value, … are uncontrollable by the designer. Finally, scenario parameters deal with future, for example economic or climate, scenarios. By ascribing these parameter categories to a different layer in a multi-layered sampling scheme as shown in FIG 1, all design options are subjected to the same uncertainties and a direct comparison for several future scenarios is enabled.

This multi-layered scheme, combined with sampling efficiency and convergence control (Janssen 2013), is proposed in Van Gelder et al. (2014) as a global probabilistic design method. In this method, first all potential design options are chosen with for example a full factorial scheme of the design parameters. Then a small multi-layered scheme is created by independent sampling of both uncertainty and scenario parameters, preferably uniformly filling the probability space. To start the Monte Carlo loop, the first design option and first scenario are selected. The small uncertainty sample is run in the model and enlarged until the desired outputs are converged. After that, the next scenarios are analogously run and more scenarios can be added until convergence of the design options or until all potential scenario values are calculated. Then, one can continue with the next design option. If all design options are converged, the outputs can be evaluated. This methodology requires execution of numerous Monte Carlo simulations, which may easily become computationally (too) expensive. The latter barrier can be overcome by use of meta-models, which mimic the original time-intensive model with a simpler and faster surrogate model.

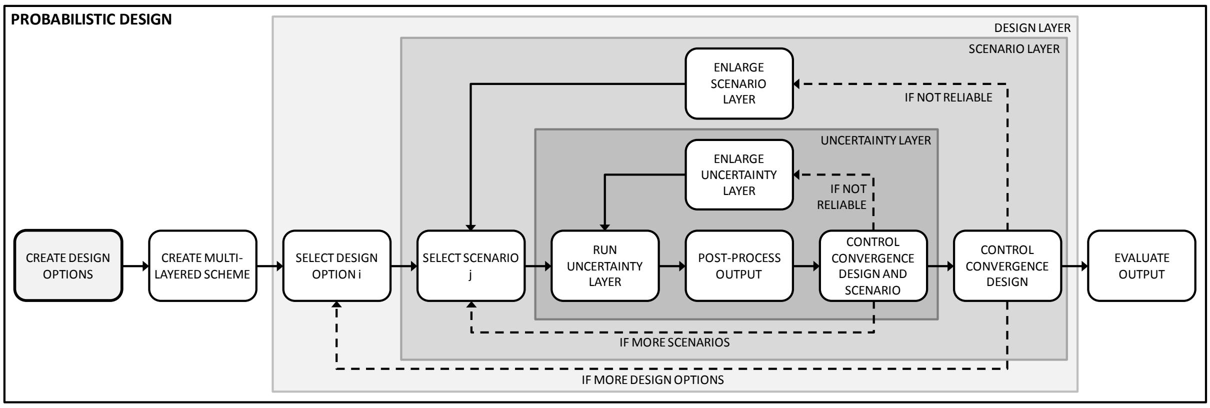

FIG 1. Probabilistic design method.

1.2 Meta-modelling

1.2.1 General aspects

Meta-models, also known as surrogate models, have the intention to mimic the original model but at a highly reduced calculation time. While for extreme cases, the original model might take days for one simulation, the meta-model only needs a fraction of this calculation time.

FIG 2 shows how to build such a meta-model based on several sample sets in order to enable cross-validation and to control calculation efficiency (Van Gelder et al. 2013b). First, all input parameters need to be sampled in a small scheme and run in the original model. Initially at least two sets are needed: one as training set to build the model, the other as validation set. Then a k-fold cross-validation is performed to control the reliability with validation indicators, which indicate how well the original model is approximated. This means that each sample set is once used as validation set, while the other sets act as training sets, resulting in as many validation indicator values as available sample sets (i.e. k). The coefficient of determination r^2, indicating the overall fit, and the maximal absolute error MAE can be used as indicators, among others. Sample sets are added until convergence of the minimal, maximal and average values of the selected validation indicators is satisfactory.

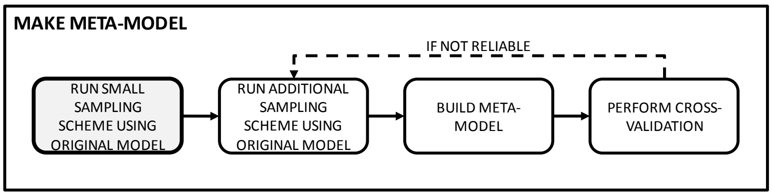

FIG 2. Meta-model construction.

1.2.2 MARS method

In this paper, cubic multivariate adaptive regression splines (MARS) (Friedman 1991, Jekabsons 2011) are used as meta-modelling method because of their good approximation ability and their fast calculation (Van Gelder et al. 2013b). Due to the use of hinge functions, model complexities can be taken into account. MARS models are of the form

$$\underline{y} = \sum_{i=1}^{m} c_i B_i(x) \tag{3}$$

Where $\underline{y}$ estimated output parameter
x input parameter vector
m amount of basis functions B_i, which can be a constant, a hinge function or a product of two hinge functions
c_i weight factors

2. Case study

To exemplify the probabilistic design method and to investigate the meta-model's reliability, a simplified case study of a semi-detached dwelling is used as shown in FIG 3. The dwelling has a floor area of 140 m², an uninsulated basement and overhangs for sun shading. Several low-energy design options are compared to select the most cost effective and robust option, with a comfortable indoor climate as auxiliary constraint. Therefore, both energy demand and maximal temperature are simulated, and net present costs are calculated afterwards.

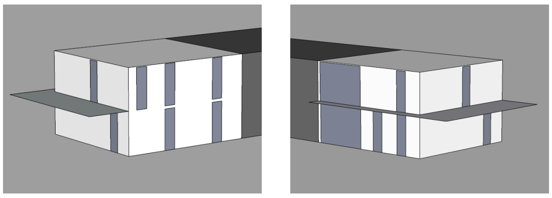

FIG 3. Dwelling model.

2.1 Output parameters

Following the European standard EN ISO 15459, the net present cost of all energy-related dwelling components is calculated over 30 years with a cost-calculation tool developed in research project IWT TETRA BEP 2020 (Verbeeck et al. 2013). The net energy demand is therefore simulated with a building energy simulation (BES) model (see section 2.3), which can be replaced by a meta-model (see section 2.4). Furthermore, the maximal temperatures in the dwelling are simulated with the BES model as well to be able to penalise those design options with a potential overheating risk.

Dwelling owners need confidence in selected design options as they require guaranteed net present costs for their investments in energy efficiency and indoor climate. Ideas from robust design are therefore incorporated by optimising mean performance and minimising spread (Zang et al. 2005).

That way, designs that best resist the uncertain and scenario parameters can be selected. Therefore, effectiveness ε and robustness R_p indicators were defined and illustrated in previous research (Van Gelder et al. 2013a). For a positive output parameter y to be minimised, the indicators are:

$$\varepsilon(x_n) = 1 - \frac{y_{50}(x_n) - y_{\min}}{y_{50} - y_{\min}} \tag{1}$$

$$R_P(x_n) = 1 - \frac{y_{50+P/2}(x_n) - y_{50-P/2}(x_n)}{y_{50+P/2} - y_{50-P/2}} \tag{2}$$

Where P user specified percentage of included sample points
y_q q^{th} percentile of distribution of y under full uncertainty
$y_q(x_n)$ q^{th} percentile of distribution of y after selecting design options x_n
y_{min} simulated minimal y value which is not an outlier, whereby an outlier is defined as a sample point smaller than y_{25}-1.5(y_{75}-y_{25})

Effectiveness ε thus describes how the deviation between median performance and optimal performance (y_{min}) improves compared to the design under full uncertainty. Robustness R_p is analogously determined as the improvement the performance spread of a design option makes in proportion to the spread under full uncertainty. According to these definitions, a solution with an effectiveness and robustness of one is the best possible, while negative values are to be avoided.

2.2 Input parameters

The design parameters considered in this paper are listed in TABLE 1. For each parameter, several

TABLE 1. Stochastic input parameters

	Parameter	Distribution*
DESIGN	Infiltration rate at 50 Pa	Dis(0.6, 1, 1.4) /h
	Ventilation system (and heat recovery)	Dis(exhaust, balanced 70% rec., balanced 80% rec.)
	U-value wall	Dis(0.1, 0.15, 0.2, 0.25) W/m²K
	Window type	Dis(1.29 W/m²K & g = 0.631, 1.31 W/m²K & g = 0.551, 0.7 W/m²K & g = 0.407)
	Sunscreen type	Dis(none, 30% transmission)
SCENARIO	Nominal energy price evolution	Dis(-1.5 %, 2.3 %, 10 %)
UNCERTAINTY	Set temperature occupancy day zone	Nor(21,1.35) °C
	Set temperature absence day zone	Dis(15°C, no reduction)
	Set temperature occupancy night zone	Nor(19,2) °C
	Internal heat gains	Uni(100,500) W
	Air change rate day zone	Wei(0.6576,4.67) /h
	night zone	Wei(1.7847,4.67) /h
	Workmanship error infiltration rate	Nor(1,0.1)
	Workmanship error U-value wall	Nor(1,0.1)
	Workmanship error heat recovery	Nor(1,0.1)

* Explanation of the symbols used:
Dis(a,b,c): discrete distribution with equal probability for a, b and c
Uni(a,b): uniform distribution between a and b
Nor(μ,σ): normal distribution wit mean value μ and standard deviation σ
Wei(λ,k): Weibull distribution with scale factor λ and shape factor k

low-energy design values are studied. To make probabilistic design with the original model feasible for this paper, calculation time is reduced by selecting only a limited set of design parameters and values. All combinations of these design values result in 216 design options.

As we are interested in the net present costs, the energy price evolution is of major interest. By considering this parameter as a scenario parameter, we are able to study the optimal results for each potential evolution. Three discrete values are considered, as shown in TABLE 1.

The inherently uncertain parameters, also listed in TABLE 1, deal with user behaviour and workmanship. The user behaviour variability is inspired by a measurement campaign of 70 new dwellings in Flanders (Belgium) (Staepels et al. 2013). 100 uncertainty layer values are sampled in sets of 20 with a *maximin* Latin Hypercube scheme (Husslage et al. 2008). In this case, this is sufficient for convergence as the maximal variation of the studied output percentiles is less than 6%. For simplicity in this paper, every design option and scenario combination is thus subjected to the same 100 samples, resulting in 64.800 simulation combinations.

Note that for clarity, in this case study, many other parameters are considered deterministic, such as occupancy profiles, climate and investment and maintenance costs.

2.3 Dynamic building energy model

The dwelling is modelled with two thermal zones and simulated in a transient BES tool developed in Modelica (Baetens et al. 2012) for the reference climate year of Uccle, Belgium (Van Gelder et al. 2013a). The adjacent dwelling is considered at a constant temperature of 19 °C. To simulate the heat demand, an ideal heating system is assumed, which is controlled using simplified occupancy and temperature profiles. A ventilation system is incorporated in the model with or without heat recovery. In summer, the heating system and heat recovery are switched off. To optimise the summer comfort, temperature of the day zone exceeds the user dependent comfort temperature, the air change rate is doubled for the next six hours or until the occupants leave the dwelling. This algorithm simulates the user behaviour to achieve a comfortable indoor climate.

2.4 Meta-models

Meta-models are built for both heat demand and maximal indoor temperature as described in section 1.2. The discrete distributions of the design parameters are transformed into uniform distributions to make other design options possible as well. All parameters are sampled together and both a sample size of 100 and 20 with up to ten sets of these sample sizes are used to build the models, as shown in FIG 4 and FIG 5. One can see that the model reliability increases with the total number of samples. Out of all presented models, four are selected to study the resulting reliability:

- reference meta-model: this is considered as the reference model as it is based on10 sets of 100 runs and is the most reliable out of the available models.
- meta-model 1: this model is built and validated on 2 sets of 100 runs and is considered as sufficiently reliable.
- meta-model 2: this model is built and validated on 10 sets of 20 runs, thus containing as many samples as meta-model 1, and the indicators are clearly converged.
- meta-model 3: this model is built and validated on 5 sets of 20 runs and is the model containing the minimal number of samples to create a reliable meta-model according to FIG 4 and FIG 5.

3. Results

As described in section 2, an optimisation is performed of the net present cost effectiveness and robustness. For that purpose, Pareto fronts are calculated. Those design options where the indoor temperature may rise above 28° C are penalised to avoid the risk on overheating. The cumulative

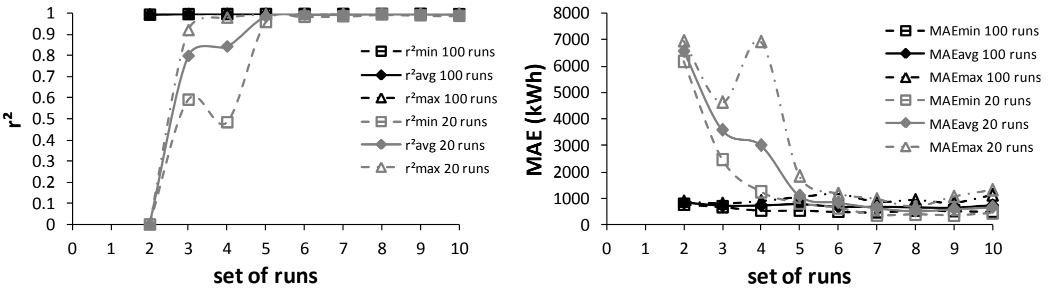

FIG 4. Minimal, average and maximal r² and MAE cross-validation indicators of the heat demand meta-model for different number of sets and samples.

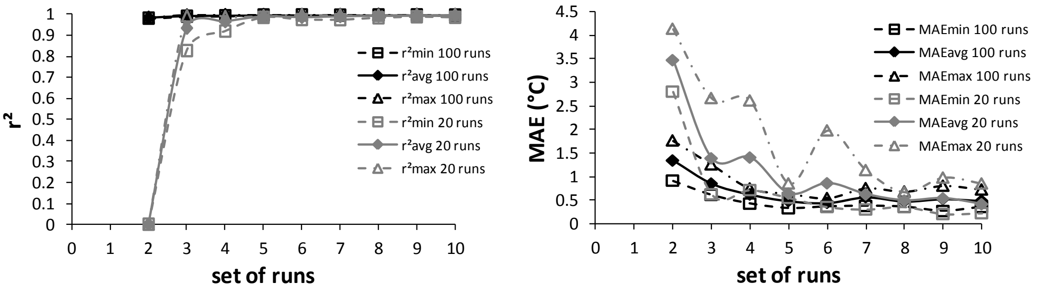

FIG 5. Minimal, average and maximal r² and MAE cross-validation indicators of the maximal indoor temperature meta-model for different number of sets and samples.

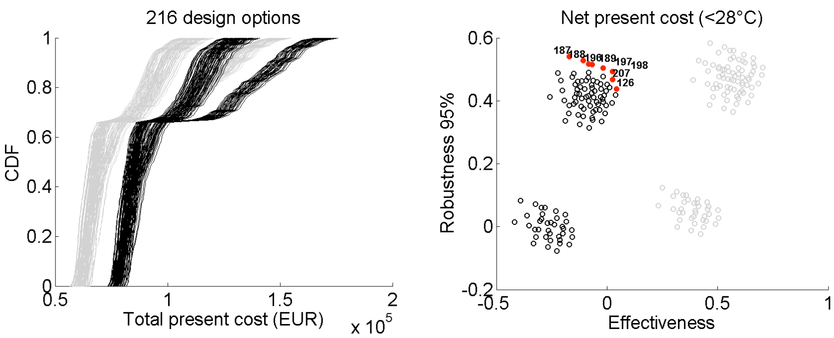

FIG 6. Cumulative distribution functions for net present cost for all 216 design options (left). Robustness R_{95} and effectiveness ε of net present cost (right). The design options with an overheating risk are indicated in grey. The Pareto front options are indicated with their design option numbers.

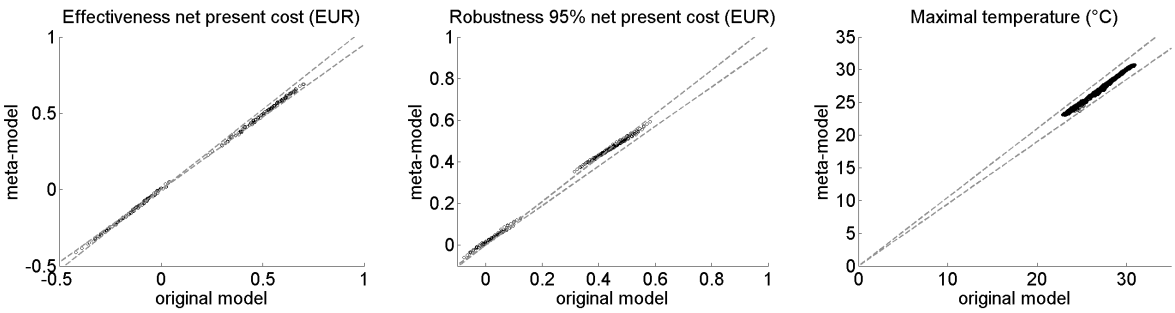

FIG 7. Comparison of outputs reference meta-model and BES model: effectiveness ε net present cost (left), robustness R_{95} net present cost (middle) and maximal indoor temperature (right). 5% deviation intervals are indicated with grey lines.

TABLE 2. Pareto front design options of dynamic BES model.

Design option n°	Infiltration rate at 50 Pa	Ventilation system (and heat recovery)	U-value wall	Window type	Sunscreen type
126	1.4 /h	balanced 80 % rec.	0.15 W/m²K	1.29 W/m²K	30 % transm.
187	0.6 /h	balanced 80 % rec.	0.10 W/m²K	0.7 W/m²K	30 % transm.
188	1 /h	balanced 80 % rec.	0.10 W/m²K	0.7 W/m²K	30 % transm.
189	1.4 /h	balanced 80 % rec.	0.10 W/m²K	0.7 W/m²K	30 % transm.
196	0.6 /h	balanced 80 % rec.	0.15 W/m²K	0.7 W/m²K	30 % transm.
197	1 /h	balanced 80 % rec.	0.15 W/m²K	0.7 W/m²K	30 % transm.
198	1.4 /h	balanced 80 % rec.	0.15 W/m²K	0.7 W/m²K	30 % transm.
207	1.4 /h	balanced 80 % rec.	0.20 W/m²K	0.7 W/m²K	30 % transm.

TABLE 3. Comparison effectiveness and robustness indicators of Pareto front design options. Grey italics indicate values which are not in the considered Pareto front.

Design option n°	BES model		Reference meta-model		Meta-model 1		Meta-model 2		Meta-model 3	
	ε	R_{95}	ε	R_{95}	ε	R_{95}	ε	R_{95}	ε	R_{95}
126	0.043	0.438	0.051	0.463	0.056	0.457	0.058	0.475	0.060	0.486
135	*0.038*	*0.418*	*0.049*	*0.445*	0.057	0.436	*0.056*	*0.451*	*0.059*	*0.464*
187	-0.171	0.540	-0.179	0.545	-0.176	0.535	-0.175	0.553	-0.177	0.556
188	-0.108	0.528	-0.113	0.536	-0.110	0.525	-0.112	0.540	-0.108	0.550
189	-0.067	0.515	-0.070	0.525	-0.067	0.515	-0.064	0.534	-0.065	0.540
196	-0.084	0.517	*-0.089*	*0.521*	*-0.087*	*0.513*	*-0.084*	*0.532*	*-0.091*	*0.533*
197	-0.018	0.505	-0.022	0.511	-0.021	0.503	-0.022	0.518	-0.022	0.526
198	0.025	0.493	0.020	0.501	0.022	0.492	0.027	0.512	0.021	0.517
207	0.025	0.467	*0.018*	*0.483*	0.023	0.471	*0.025*	*0.489*	*0.021*	*0.494*

distribution functions (CDF) of all design options, needed to calculate ε and R_{95}, and the Pareto front options of the BES-model optimisation are shown in FIG 6 and listed in TABLE 2.

When comparing net present cost effectiveness ε and robustness R_{95} and maximal indoor temperatures between BES model and reference meta-model, slightly deviating values are found, as shown in FIG 7. Although these deviations become slightly larger when fewer samples are used to build the meta-model, very similar Pareto fronts are obtained, as presented in TABLE 3. Only one option (i.e. 135) appears that was not in the original Pareto front. But this design option is very similar to option 126, as only the U-value changes (0.2 W/m²K). On the other hand, options 196 and 207 do not appear in the meta-model Pareto front, but they are almost equal to the other options and are still close to the Pareto front. Note that the optimal ε values are very small due to the fact that most effective solutions result in overheating risks.

Similar observations remain when comparing Pareto fronts per scenario. Those results are not explicitly presented here. When comparing Pareto fronts from meta-models built on fewer samples than meta-model 3, larger deviations are found. Moreover, design options with an overheating potential might be selected as this risk is unreliably detected. FIG 4 and FIG 5 show that these meta-models are indeed less reliable as they have low r^2 values and large maximal errors.

4. Conclusions

As illustrated in section 3, meta-models can be reliably used in probabilistic design of low-energy dwellings as output distributions and effectiveness and robustness are sufficiently mimicked and very similar Pareto optimal design options are found. This allows performing a generally time-consuming probabilistic design as presented in section 1.1, but now in only a fraction of the original time. The presented method uses multi-layered schemes to classify parameters by their physical meaning as this enables the comparison of numerous design options and scenarios.

In order to reliably and time efficiently build a meta-model, a model procedure based on replicated sample schemes was proposed in section 1.2. Small schemes are preferred as it is seen that meta-model build on those schemes perform as well as the others, but less samples are needed.

5. Acknowledgements

The authors are very thankful for the funding of the Flemish government and companies for the IWT TETRA BEP2020 project (Verbeeck et al. 2013) and participating inhabitants for the related measurement campaign. Many thanks to Liesbeth Staepels to develop the cost-calculation tool according to the European standard EN ISO 15429. They would like to thank Bart Husslage and Gijs Rennen from the Tilburg University as well for sharing their MATLAB code for calculation of maximin sampling schemes (Husslage et al. 2008).

References

Baetens, R., De Coninck, R., Van Roy, J., Verbruggen, B., Driesen, J., Helsen, L. & Saelens, D. 2012. Assessing electrical bottlenecks at feeder level for residential net zero-energy buildings by integrated system simulation. Applied Energy, 96:74–83.

EN ISO 15459. Energy performance of buildings – Economic evaluation procedure for energy systems in buildings.

Friedman, J.H., 1991. Multivariate adaptive regression splines. The Annals of Statistics, 19(1), pp.1–141.

Husslage, B. et al., 2008. Space-filling Latin hypercube designs for computer experiments.

Janssen, H., 2013. Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence. Reliability Engineering & System Safety, 109:123–132.

Jekabsons G., 2011. ARESLab: Adaptive Regression Splines toolbox for Matlab/Octave, available at http://www.cs.rtu.lv/jekabsons/

Staepels, L., Verbeeck, G., Roels, S., Van Gelder, L. & Bauwens, G., 2013. Evaluation of Indoor Climate in Low Energy Houses. In Symposium on Simulation for Architecture and Urban Design, April 7-10. San Diego, USA.

Van Gelder, L., Janssen, H., Roels, S., Verbeeck, G. & Staepels, L., 2013a. Effective and robust measures for energy efficient dwellings : probabilistic determination. In Building Simulation 2013. Chambéry, France, pp. 3465–3472.

Van Gelder, L., Janssen, H. & Roels, S., 2013b. Metamodelling in robust low-energy dwelling design. In A. Mahdavi & B. Martens, eds. 2nd Central European Symposium on Building Physics, September 9-11. Vienna, Austria, pp. 93–99.

Van Gelder, L., Janssen, H. & Roels, S., 2014. Probabilistic design and analysis of building performances: methodology and application example. Energy and Buildings. Under review.

Verbeeck, G., Staepels, L., Roels, S., Van Gelder, L., Bauwens, G. & Deconinck, A.-H., 2013, IWT TETRA BEP2020: betrouwbare energieprestaties van woningen – Naar een robuuste en gebruikersonafhankelijke performantie (in Dutch), URL http://bep2020.pxl.be.

Zang, C., Friswell, M.I. & Mottershead, J.E., 2005. A review of robust optimal design and its application in dynamics. Computers & Structures, 83(4-5), pp.315–326.